

Integrated Testlets: A New Form of Expert-Student Collaborative Testing



2015 CELT Vol. VIII
celt.uwindsor.ca
www.stlhe.ca



Ralph C. Shiell and Aaron D. Slepkov
Trent University

Integrated testlets are a new assessment tool that encompass the procedural benefits of multiple-choice testing, the pedagogical advantages of free-response-based tests, and the collaborative aspects of a viva voce or defence examination format. The result is a robust assessment tool that provides a significant formative aspect for students. Integrated testlets utilize an answer-until-correct response format within a scaffolded set of multiple-choice items that each provide immediate confirmatory or corrective feedback while also allowing for the granting of partial credit. We posit here that this testing format comprises a form of expert-student collaboration, we expand on this significance and discuss possible extensions to the approach.

Introduction

Course assessment is a key component of university courses, and yet in comparison to the delivery of course content, the methodology of assessment is less frequently considered. It is rare to reflect why and how we assess, and even more rare to address how to most effectively conduct the assessment (Mazur, 2013). The most immediate purposes of classroom tests are both to assess students' learning outcomes and to act as a motivator for students (Ebel & Frisbie, 1991), yet instructors now increasingly have many additional objectives from conducting assessments, including providing *formative* experiences such as practice in problem solving, opportunities for meta-cognitive reflection, and confidence-boosting opportunities.

Even within a purely summative context, it has long been assumed that assessment through a set of free-response questions (also called *constructed-response* questions) is the most effective approach to assess student understanding. Here a student

generates an acceptable response by demonstrating their integration of a wide and often complex set of skills and concepts. To score the question, an expert interprets each response and gauges its level of "correctness." In contrast to these are multiple-choice questions (termed *items*), where response options are provided with the correct answer (the *keyed option*) listed along with several incorrect answers (the *distractors*). The student's task is then to select the keyed option from this list. Free-response questions are usually presumed a more valid assessment tool as they do not provide students with the correct answer and are perceived to better assess the combination of cognitive processes needed for solving problems that integrate several concepts and procedures. The explicit solution synthesis required by free-response questions furthermore suggests to instructors a strong (but often false) sense of transparency of student thinking. Nonetheless, the scoring of multiple-choice items is quicker, more reliable and cheaper (Haladyna, 2004), and with proper construction, multiple-choice items can be powerful tools for the

assessment of conceptual knowledge (DiBattista, 2008). Many introductory final exams consist entirely of multiple-choice questions, where the procedural advantages of multiple-choice testing are weighed against any pedagogical disadvantages stemming from an exam format that may necessarily measure only compartmentalized conceptual knowledge and calculation procedures. Overall the use of a multiple-choice format for formal assessments in many disciplines is not wholeheartedly embraced, and, when possible, greater exam weight is still typically placed on traditional free-response questions that require explicit synthesis to solve the problem at hand.

To address the perceived drawbacks of multiple-choice testing a number of variants have been introduced; specifically in order to assess complex cognitive processes and/or to reward partial knowledge. These include manipulating the choices given to students so that options contain different combinations of primary responses only some of which are true (complex multiple choice, type K, true-false or type X, and multiple-response formats) (Berk, 1996), manipulating the stems by asking students for predictive or evaluative assessments of a scenario rather than simply recounting knowledge (Berk, 1996), confidence or probability weighting of options (Ben-Simon, Budescu, & Nevo, 1997), and the “multiple response format” in which multiple stages are created within each multiple-choice item, with scores weighted according to whether the reasoning is correct (Wilcox & Pollock, 2014). Interpretive exercises consist of a series of items based on a common set of information/data/tables, with each item requiring students to demonstrate a particular interpretive skill to be measured (Linn & Miller, 2005). Assessment goals such as recognizing assumptions, inferences, conclusions, relationships and applications can each be independently measured. Meanwhile, another framework of assessment, *collaborative testing*, that specifically addresses formative goals is rapidly gaining in popularity. Here students initially write a test as individuals and then form small groups to rewrite the test, with consensus required for each response before submission. The marks are typically weighted for the

two stages 85%:15% respectively, and such testing brings both formative and meta-cognitive aspects to the assessment, with increased learning taking place under such a setting (Gilley & Clarkston, 2014). Many of the advantages of collaborative testing, including knowledge gain, are believed to result from the dialogue between students and their peers (Wieman, Rieger, & Heiner, 2014).

We have recently invented a new multiple-choice-based assessment platform that is designed to combine the procedural advantages of multiple-choice testing with the pedagogical advantages of free-response, while also contributing to a formative nature of assessment. Such *integrated testlets* (ITs) utilize an answer-until-correct response format within a scaffolded set of multiple-choice items that each provide immediate confirmatory or corrective feedback while also allowing for the granting of partial credit. We posit that the skilful engineering of question scaffolding together with the anticipation of students receiving immediate feedback during the test comprises a form of passive expert-student collaboration. In this article first we introduce ITs and then describe their construction and operation, specifically exploring the notion that they embody aspects of collaborative testing.

Integrated Testlets

While conventional *testlets* (Haladyna, 1992) and interpretive exercises are multiple-choice item sets with a common context but composed of independent items, an IT purposefully interrelates the multiple-choice items so that knowledge of the answer for a given item is helpful or even required for answering subsequent items. The degree to which solving later items depends on the answers from former items defines the extent of integration in an IT. We typically denote ITs as either “weakly-integrated”, “moderately integrated”, or “strongly integrated”, while traditional testlets would be considered “non-integrated”. Adopting an answer-until-correct approach permits our deployment of such an integrated set of multiple-choice items because it avoids a ‘double jeopardy’ situation (where

a student is unknowingly penalized twice; once for an initial item which is answered incorrectly, and again in a subsequent item which requires this previous answer), and it also permits all students, regardless of their score on earlier items, to progress through the testlet. The correct answer to each item is conveyed to the students with full/partial/zero marks awarded as appropriate before they proceed to the next item with full knowledge of the correct answer. For our particular implementation of this approach we choose to use commercially available Immediate Feedback Assessment Technique (IF-AT) cards (Epstein et al., 2002) with boxes coated in a similar way to scratch-and-win lottery tickets, concealing a star within the keyed-response option and the distractor options being blank. Students answer each item until a star is revealed, and they then advance to the next item within the testlet with full knowledge of the answers to all previous items. In addition to being able to access higher-level learning, students also leave the exam with full knowledge of their score. It has been demonstrated that such an answer-until-correct approach is substantially preferred by students compared to the “Scantron” method (DiBattista, Mitterer, & Gosse, 2004). Moreover, immediate feedback has been demonstrated to improve learning outcomes relative to the results observed with delayed feedback (Dihoff, Brosvic, Epstein, & Cook, 2004).

Figure 1 shows a research-validated integrated testlet (Slepkov & Shiell, 2014) specifically designed to test higher-level thinking in a first-year Introductory Physics course. The topic is that of mechanics, and involves the understanding and application of the vector nature of forces, determining friction, Newton’s second law, and one- and two-dimensional kinematics, with items aligned to particular learning outcomes of the course. It is an example of a strongly-integrated testlet, as will be described below. This particular IT was designed to replace a free-response question and therefore aims to test analytical, conceptual, evaluative, and procedural knowledge. For the purposes of this article, we do not presume the reader to have an understanding of the physics needed to solve the IT, nor do we aim to teach such knowledge here. Rather, we use this testlet as a

canonical example of the construction and operation of ITs.

As part of a formal comparison between IT and free-response formats in exams within an Introductory Physics class we deployed a set of concept-equivalent ITs and free-response questions and found both formats to be both highly discriminating and reliable (Slepkov & Shiell, 2014). A purely psychometrics-based analysis suggested that the free-response format was marginally better at both these measures, but further analysis exposed a large inter-rater variability with the free-response format scoring, while also suggesting that the range of marks awarded for free-response was artificially dispersed, with students between the top and bottom cohorts receiving scores that were only weakly proportional to their mastery of the material. Some additional advantages of ITs are the reduced time it takes students to complete a question, and that the resulting grade distributions appear to more reliably reflect students’ knowledge. Overall we find that ITs are a highly-effective multiple-choice testing platform for assessing deeper knowledge.

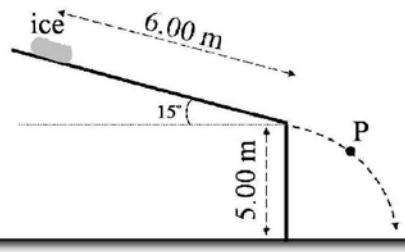
To date we have composed approximately forty ITs in physics (our principal discipline), ten ITs in chemistry, and single ITs in each of calculus, biology, psychology, art history, and 20th century literature. These are scaffolded and integrated to different extents, with the strength of integration roughly scaling with the quantitative nature of each discipline. We now summarize how we design and deploy ITs, specifically with reference to the example given in Figure 1, and further we make a case for how ITs can embody a collaborative conversation between instructor and students.

Construction and Implementation of Integrated Testlets: a Collaborative Conversation

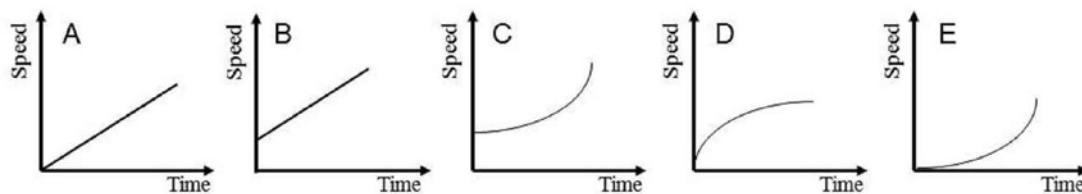
The first step in composing an integrated testlet is the identification of a complex problem. In the introductory physics example shown in Figure 1 the

Kinematics and projectile motion

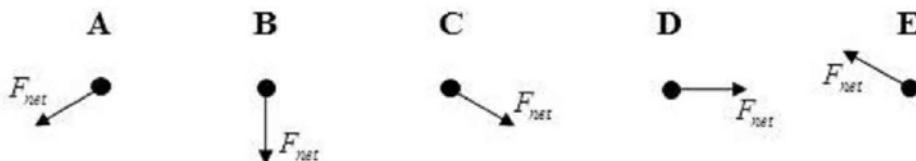
A piece of ice is 6.00 m from the edge of a roof of height 5.00 m at an angle of 15° to the horizontal as shown in the figure, when it begins to slide. The coefficient of kinetic friction between the ice and the roof is 0.2. Ignore any air resistance.



- 1) Which of the following graphs best represents the speed of the ice as a function of time, beginning from the moment when it first begins to slide and ending when it leaves the roof?



- 2) Which of the following free-body-diagrams is most correct for the ice at position P?



- 3) How much time elapses between the ice beginning to slide and it leaving the roof?

A. 0.65 s B. 4.3 s C. 1.1 s D. 2.2 s E. 1.6 s

- 4) At what (horizontal) distance from the base of the house does the ice land?

A. 5.3 m B. 1.9 m C. 19 m D. 3.5 m E. 2.5 m

Ans: A:B:B:E

Figure 1

An example of a strongly-integrated testlet from an Introductory Physics course. This particular IT tests mechanics, and specifically the vector nature of forces, Newton's 2nd law, projectile motion, and kinematics.

problem to be solved is determining how far away from the side of a house a piece of ice lands after it slides off a roof. Such a question is a mainstay of traditional free-response exams and homework assignments, but is typically too complex for assessment by multiple-choice items. Our ITs usually (but not always) consist of four multiple-choice items, each representing a non-trivial step in solving the problem. The concepts and procedures for solving the problem are deconstructed much as one would when composing a scoring rubric and each multiple-choice item often increasingly and cumulatively mines students' abilities within the cognitive process dimension and/or the knowledge dimension of the revised Bloom's taxonomy (Anderson & Krathwohl, 2001). In fact, we often "reverse-engineer" our ITs by formally solving a targeted free-response problem, constructing a scoring rubric that is based on our assessment/learning objectives, and then reconstructing a set of multiple-choice items that span these objectives. The actual choice of multiple-choice items depends on many considerations such as the size of the procedural or cognitive leap between items, the extent of the requirement of the knowledge from any given intermediate step to cue the next step, and the importance of any intermediate step to fulfil our learning objectives. As a concrete example, Item 1 in Figure 1 assesses the student's ability to resolve forces into their components, to apply knowledge that kinetic friction exerts a constant force on a moving object, and finally to apply Newton's second law to determine that the acceleration of the ice down the roof is constant. Thus, Item 1 is already more than a simple recollection or identification-based multiple-choice question. Item 2 requires students to appreciate the origin of forces, and to determine that despite the fact that the ice travels along a curved path in the air it does so while being acted on by a single constant force (gravity). Item 3 then requires the application of the kinematic distance-time equation to an object experiencing the motion represented in Item 1. Finally, Item 4 extends this to the case of two-dimensional motion. Thus, Item 4 is highly scaffolded by Items 1, 2, and 3. In fact, as described below via "integration maps", the solution to Item 4

unquestionably depends on the solutions to Items 2 and 3.

For ITs to work well, we closely follow the best practices for multiple-choice question construction (Frey, Petersen, Edwards, Pedrotti, & Peyton, 2005). Thus the *stimulus*, i.e. the initial text and diagram that describe the problem within an IT, is clear and consistent, containing as much information as possible while avoiding irrelevant details. The IT in Figure 1, as is often the case, also contains a diagram as part of its stimulus. Consistency in wording is particularly important. For example, after initially introducing the ice, it is referred to using the same nomenclature within all items that comprise the IT. The stems within all items are then formally written as questions. Furthermore diagrams are labelled unambiguously. Note for example that the designated point on the trajectory is unambiguously labelled with a "P", rather than "A", which is instead used as an option label, or a "1", which can be confused with a magnitude of some sort.

Much consideration goes into the construction and choice of distractors, particularly in an integrated testlet where corrective feedback can actively address major student misconceptions and thus supports how an item provides scaffolding for subsequent items. For each item the distractors are constructed by anticipating students' answers to the item, and often involve knowing common misunderstandings in either concept or application. For example, in Item 1 the two distractors C and E present the misconception that a constant net force acts to increase speed (true) but in a nonuniform/nonlinear way (false). Distractor D presents the misconception that a terminal speed is reached, which is not valid in this situation. Both distractors B and C present the misconception that the speed starts with an offset, or "trapping" students to conflate a starting height offset with a starting speed offset. Similarly, in Item 2 the distractors correspond to the commonly-encountered confusion concerning the direction of the acceleration of an object when it is already moving. Items 3 and 4 are numerical questions, for which there are an infinite number of possible incorrect answers. Here we present numerically-plausible distractors, some of

which are derived from typical mathematical errors or mathematical misconceptions. Furthermore, the precise choice of numerical distractors is considerate of a common student practice of “edge avoidance” (i.e. we often allow the keyed response to be the highest or lowest available value).

What makes ITs unique is that while answering each item students have a form of passive conversation with the instructor after they make their response, and they then either continue (if they have in fact chosen the keyed option) or they pause and refine their thinking (if they have chosen a distractor). This is a two-way conversation: The instructor has choices in how they scaffold the questions, the extent to which they wish to cue certain concepts, and their choice of distractors. For example, depending on the assessment goals of the instructor, they may choose to exclude a distractor that represents a simple and non-instructive “trap”. Thus, if a student arrives at such an answer (for example, due to a trivial mistake) they find it absent and thus revisit their thinking. In effect, the instructor’s anticipation of such a mistake—and their avoidance of trapping for it—is part of the conversation with the student. Likewise, deliberately choosing to include a distractor that traps for a key misunderstanding is also part of the conversation: the student discovers that they have made an error and by subsequent selection of a correct response has been informed (in effect by the instructor) that their original thinking was flawed. This conversational interpretation of student thinking is supported by an analysis of the partial marks awarded in ITs, which themselves were found to be highly discriminating (Slepkov & Schiell, 2014). That is, those students in the upper quartiles earned a higher fraction of available partial marks than those in lower quartiles, which implies that students improve their understanding in a selective and proportionate manner. To be sure, such a delayed passive conversation does not represent fully active peer-student and expert-student collaboration, but it does share some of the immediate-feedback attributes of collaborative testing. Unlike peer collaborative testing, however, with ITs the student always ends up with the “expert” answer (i.e. the correct answer). Thus, we view ITs as passive expert-student

collaborative tests, and we shall conduct further studies, involving analyzing time-sequences of students’ rough written work and post-test interviews to more concretely determine the validity of this perspective.

There is some preliminary evidence that such a conversation takes place during IT-based examinations. In our previous study (Slepkov & Shiell, 2014) we surveyed students after an Introductory Physics midterm exam that contained two ITs and two free-response questions, each of which covered independent and different topics. We asked students “For the multiple-choice parts of the midterm (i.e. the testlets), did you use answers you uncovered from the early questions to answer any of the later questions in a testlet?” A substantial 90% of the students said they had done so at least once. This indicates that most used the scaffolding, and therefore the implicit conversation described above, as we had intended. On the other hand, while scoring the free-response questions it became evident that if students were confused or ignorant about how to begin to answer the question, they had very few tools to allow them to demonstrate partial knowledge or how to answer the rest of the question. The lack of scaffolding opportunities within the free-response format is a major disadvantage of that technique over ITs.

As part of IT design, we find the creation of *integration maps* to be a highly useful endeavour. *Integration maps* represent for the instructor the flow of cognitive processes involved in moving through an IT, which themselves can be represented by a *concept map*, as shown in Figure 2a. This shows the individual steps involved in working through the complete problem from stimulus to answering the last item. The integration map, shown in Figure 2(b), can then help the instructor to select particular items for the IT. This map makes clear the relationships between Items (questions) 1-4. As mentioned above, the solution to Item 1 only weakly informs the answering of Item 3, whereas the solutions to Items 2 and 3 are required to obtain the solution to Item 4. Item 1 and Item 2 are independent, but together they aid to scaffold Item 4. The opportunity to grant partial credit in a multiple-choice exam is a major boon to

the IT approach. The IF-AT cards, for example, allow this by simply assigning marks based on the number of tries a student took before uncovering the correct response. The precise choice of marking scheme for items, and therefore the proportion of partial credit granted, will affect both how students approach each IT and influence the test psychometrics (such as mean test score and measures of item discrimination). In five-option items, we typically grant full marks for the selection of a correct answer in the first response, half-marks for correct responses in the second selection, and one-tenth-marks for correct responses in third selections; with no marks given for subsequent

selections. This scoring scheme, designated [1, 0.5, 0.1, 0, 0], has been adopted as a balance between keeping the expectation value for guessing sufficiently low as to make passing of the test statistically unlikely due to guessing alone with a desire to prolong students' intellectual engagement with items via partial credit incentives.

We could however use any number of alternate schemes. We have assessed the effects of marking students based on a variety of schemes ranging from "generous" (i.e. [1,0.7,0.3,0,0]), through "harsh" (i.e. [1,0.3,0.0,0]) to "dichotomous" (i.e. [1,0,0,0,0]).

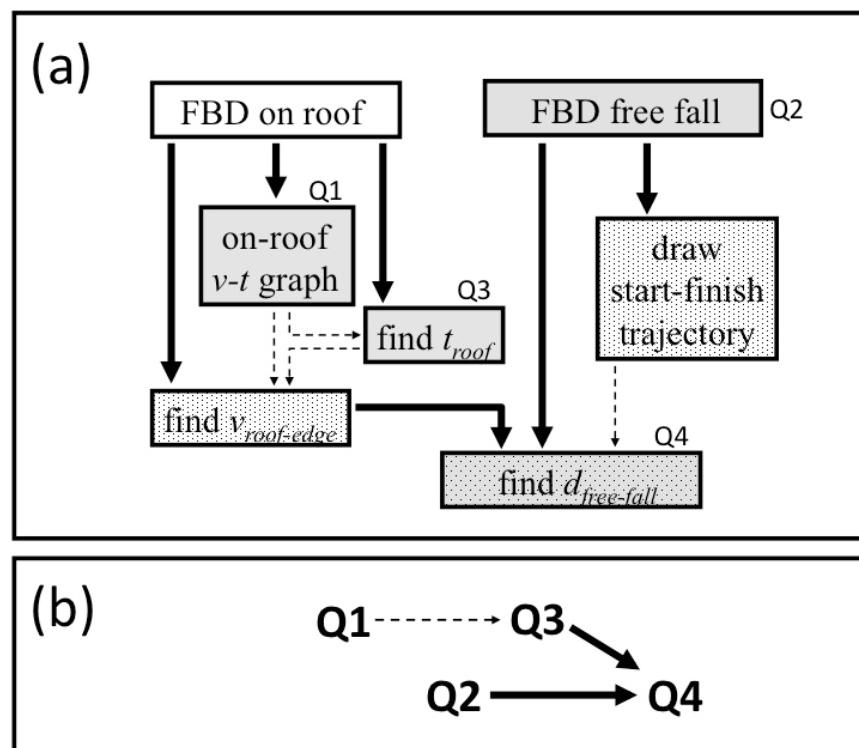


Figure 2

(a) A concept map for the integrated testlet shown in Figure 1. Items 1-4 are labelled Q1-Q4 and the steps between them are shown with arrows. (b) An integration map that summarizes the interrelationship between the items chosen for this IT. A dashed arrow indicates a weak relationship between questions, where knowledge of one item may potentially inform the other, while a solid arrow indicates that the solution of the latter question is predicated on knowledge of the former. In this case, explicit knowledge of the concepts/procedures tested in Q2 and Q3 are needed for the solution of Q4.

We find that compared to dichotomous scoring, adding any reasonable partial credit scheme slightly boosts the average score, while maintaining the highly-discriminating nature of a good test (Slepkov & Shiell, 2014). Certainly, we have seen no evidence that partial credit dilutes the discriminatory power of the test.

The gold-standard of testing—albeit impractical in a classroom setting—is through a *viva voce*, or oral defence, format. Such an examination truly represents an active expert-student collaborative test. Further supplementing the perspective of ITs as an (albeit delayed) collaborative conversation between expert and student are the other ways in which ITs can closely share the benefits of a *viva voce* format, but which are absent in both multiple-choice- and free-response-based exams. One example within a quantitative discipline such as physics is to ask students to recall (or determine) from a list the correct representation of a formula that can usually be found on their formula sheet, but is redacted in this circumstance. This formula can then be used within subsequent items in the IT. By composing distractors in the manner described above, the expert engages in a “delayed-discussion” with the students and, further, provides expert guidance during the assessment should a student not initially select the keyed option, which in this case corresponds to the correct version of the formula. This is very similar to a dialogue that frequently occurs within an oral examination, where the student is first probed on fundamental laws in science (i.e. the relevant equations), before these are then applied to the particular problem at hand.

Conclusions

An integrated testlet (IT) is a relatively new assessment tool that measures students' understanding of complex ideas through a set of scaffolded multiple-choice items, each adopting an answer-until-correct format. Students continue answering each item within an IT until the correct

answer is revealed to them, and they then advance to the next item with full knowledge of, and benefit from, answers to previous items. ITs can be valid and efficient replacements for free-response questions, as they assess complex cognitive processes and can also reward partial knowledge. We posit that this testing format comprises a form of expert-student collaboration, approaching the gold-standard of a *viva voce*, or oral defence, format. The extent of the delayed-discussion between expert and student has been discussed, reflecting the expert guidance given during the assessment to those students who do not initially select the keyed option for an item within an IT. Indeed, ITs in scientific disciplines may be adapted to even better replicate an oral examination by first building up the foundational principles underpinning particular concepts, and then, after that “conversation” is concluded successfully, subsequently apply these concepts to a real-world situation that is almost always too complicated for a stand-alone multiple-choice or free-response approach. This would constitute a super-IT or an integrated (interdependent) set of ITs. An entire exam could then comprise a flowing set of related testlets, with immediate confirmatory or corrective feedback at each step — a significant leap towards that which happens in a *viva voce* exam but with the reliability and streamlined advantages of multiple-choice testing.

References

- Anderson, L.W., & Krathwohl, D.R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Ben-Simon, A., Budescu, D. V., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*,

- 21, 65-88. <http://dx.doi.org/10.1177/0146621697211006>
- Berk, R. A. (1996). A consumer's guide to multiple-choice item formats that measure complex cognitive outcomes. In *National evaluation systems, from policy to practice* (pp. 101–127). Amherst, MA: Pearson Education. Retrieved from http://images.pearsonassessments.com/images/NES_Publications/1996_12Berk_368_1.pdf
- DiBattista, D. (2008). Making the most of multiple-choice questions: Getting beyond remembering. *Collected Essays on Teaching and Learning*, 1, 119-122. <http://celt.uwindsor.ca/ojs/leddy/index.php/CELT/index>
- DiBattista, D., Mitterer, J. O., & Gosse, L. (2004). Acceptance by undergraduates of the immediate feedback assessment technique for multiple-choice testing. *Teaching in Higher Education*, 9, 17-28. <http://dx.doi.org/10.1080/1356251032000155803>
- Dihoff, R. E., Brosvic, G.M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *The Psychological Record*, 54(2), 207-231. <http://opensiuc.lib.siu.edu/tpr/vol54/iss2/4>
- Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of Educational Measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, 52, 187-201. <http://thepsychologicalrecord.siu.edu>
- Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education*, 21, 357-364. <http://dx.doi.org/10.1016/j.tate.2005.01.008>
- Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching*, 43(3), 83-91. http://learningcenter.nsta.org/browse_journals.aspx?journal=jcst
- Haladyna, T. M. (1992). Context dependent item sets. *Educational Measurement: Issues and Practice*, 11(1), 21-25. <http://dx.doi.org/10.1111/j.1745-3992.1992.tb00223.x>
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Assoc.
- Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, NJ: Pearson.
- Mazur, E. [Derek Bok Center, Harvard University]. (2013, Nov 19). *Assessment: The silent killer of learning* [Video file]. Retrieved from <https://www.youtube.com/watch?v=CBzn9RAJG6Q>
- Slepkov, A.D., & Shiell, R. C. (2014). Comparison of integrated testlet and constructed-response question formats. *Physical Review Special Topics - Physics Education Research*, 10(2), 020120-1 - 020120-15. <http://dx.doi.org/10.1103/PhysRevSTPE.R.10.020120>
- Wieman, C. E., Rieger, G. W. & Heiner C. E. (2014). Physics exams that promote collaborative learning. *The Physics Teacher*,

52, 51-53. <http://dx.doi.org/10.1119/1.4849159>

Wilcox, B. R., & Pollock, S. J. (2014). Coupled multiple-response versus free-response conceptual assessment: An example from upper-division physics. *Physical Review Special Topics - Physics Education Research*, 10(2), 020124-1 – 020124-11. <http://dx.doi.org/10.1103/PhysRevSTPE.R.10.020124>

Acknowledgements

We thank David DiBattista of Brock University for guidance and advice, and Rachel Wortis and Bill Atkinson of Trent University for useful discussion and encouragement.

Biographies

Ralph Shiell is Associate Professor of Physics at Trent University where he is engaged in pedagogy research and optical physics research, and has taught at all undergraduate and graduate levels since 2004. His interests in physics education come from an appreciation of science as a highly-connected entity, with distinct concepts that together enable a deep understanding of the natural world.

Aaron Slepkov is Assistant Professor and Canada Research Chair in the Physics of Biomaterials who is haunted by the question of why physics is so unpopular among students. En route to a satisfactory answer, Aaron works to develop and measure innovative and effective assessment tools for science education.